1993

1. Curriculum-Based Measurement

Stanley L. Deno
University of Minnesota, denox001@umn.edu

1

Curriculum-Based Measurement

Stanley L. Deno

University of Minnesota

In the history of Greek mythology there is a character named Sisyphus who, for sins committed during his lifetime, is condemned to spend eternity pushing a boulder up a hill. No matter how hard Sisyphus tries as he nears the top of the hill, the boulder rolls back down. Sisyphus cannot escape from this continued cycle of effort and failure. Sometimes, when I think about the experiences of many children attempting to learn basic skills in the public schools, I think of the myth of Sisyphus. Too often, it seems to me, no matter how hard they try, they do not succeed.

In this chapter on curriculum-based measurement (CBM) I want to focus on three points: First, what is curriculum-based measurement? Second, why was curriculum-based measurement developed? And third, how does the use of curriculum-based measurement help to avoid the problem of Sisyphus and education?

CURRICULUM-BASED MEASUREMENT DEFINED

As a Subset of Curriculum-Based Assessment

The term *curriculum-based assessment* (CBA) is a very popular topic in the field of special education these days (Tucker, 1985). As Tucker points out, CBA is a term used to describe a practice that has existed for a very long time—the practice of using what is to be learned as the basis for assessing what has been learned. Since traditional psychometric test construction also involves use of the table of specifications to define the content domains for which test items must be developed, the difference between CBA and traditional psychometric testing may not be immediately obvious. However, four salient differences between CBA and traditional psychometric testing can be identified: First, in CBA, the very curriculum materials that serve as the media for instruction are used as the test stimuli; second, direct observation and recording of student performance in response to selected curriculum materials are emphasized as a basis for collecting the information used to make assessment decisions; third, interobserver agreement is the primary technique used to establish the reliability of information collected through CBA; and fourth, social validity is typically the basis for justifying the use of information gathered through CBA. Given these emphases, it is common for CBA proponents to argue that the information gathered from student performance in the curriculum more adequately reflects the real goals of instruction in the classroom than most standardized achievement tests, because the assessment information obtained through CBA relates more directly to what is being taught, and also because the content and materials of daily instruction are a fairer and firmer basis for making judgments about student learning.

Since the focus here is on CBM, some clarification of the term is needed. The term *assessment* as used in CBA is a very broad term that refers to information gathered for purposes of decision making. Thus, curriculum-based assessment refers to all sorts of information-gathering practices that may occur when observing student performance in the curriculum. These practices include scoring the student's worksheets to obtain a percentage score for the problems or answers correctly completed on a worksheet; making judgments about a student's reading comprehension based on the prosodic features of that student's oral reading; and moving the student to a new skill based on consecutive days of answering all questions correctly. In CBA, typically, different assessment information is collected for different decisions. A variety of different but related approaches to CBA are represented in the current literature (cf. Howell & Morehead, 1987; Bigge, 1988; Idol, Nevin, & Paolucci-Whitcomb, 1986; Shinn, 1989).

As Distinct from CBA

Curriculum-based measurement (CBM) is a separate and distinct subset of CBA procedures, a specific set of steps for measuring student growth in basic skills, developed at the University of Minnesota through the Institute for Research on Learning Disabilities (IRLD) (Deno, 1985). The CBM procedures were developed as part of a larger program of research directed toward designing a practically feasible and effective formative evaluation system that special education teachers could use to build more effective instructional programs for their students. As part of that formative evaluation system, it was necessary to create a simple, reliable, and valid set of measurement procedures that teachers could use to measure frequently and repeatedly the growth of their students in the basic skills of reading, spelling, and written expression. When these procedures are used within the context of the local school's curriculum, they become CBM.

Like CBA, in general, CBM focuses on using existing curriculum materials and goals as a basis for selecting and creating the tasks on which student performance is measured. The primary difference is that CBM is more limited with respect to quantification procedures and types of information collected than is the case with CBA. The term *measurement* in CBM is used to denote the focus on the use of standardization to produce a technically adequate quantitative scale—an issue of less concern in most other CBA models. Although differing in some respects, all curriculum-based approaches share the assumption that data collected from observations of day-to-day student performance in the curriculum are, at the very least, an important supplement for making a broad range of educational decisions. Indeed, a basic assumption is that curriculum-based approaches may be a necessary alternative to commercially distributed achievement tests if measurement is ever going to contribute to educational improvements. Also, curriculum-based advocates generally share the view that traditional approaches to assessment and measurement have failed to contribute sufficiently to educational improvement and that alternatives, such as curriculum-based approaches, offer greater promise.

An Example of CBM

The set of measurement procedures referred to here as CBM were developed through the University of Minnesota IRLD during the years of 1977—83. These research and development activities focused on creating measurement procedures for clearly and simply describing growth in functional literacy. Subsequently, school districts have used similar approaches to develop measures of basic numeracy. Since the focus of all of these research and development activities has been on students who were having significant difficulty developing literacy and numeracy, most (but not all) of the work has been with students in elementary and middle schools. In Figure 1, an illustration of the results of using CBM procedures with a student in reading over the course of a school year is displayed. As can be seen, student performance in terms of the number of words read aloud correctly in 1 minute from the student's grade level basal reader is presented simply and clearly in relation to changes made in that student's instruction. Although the graph is interesting, the question that needs to be addressed is why so much time and energy were spent to produce such a graph.

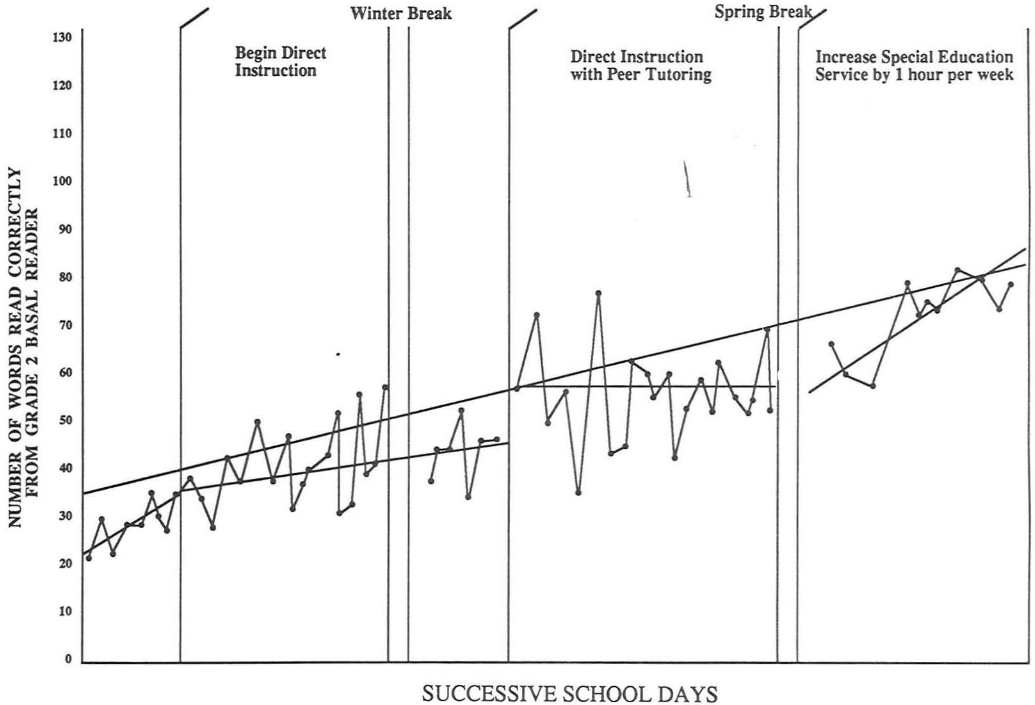
WHY CBM?

A Brief Personal History

In the early 1970s at the University of Minnesota, we were attempting to develop a field practicum site that the Special Education Program could use for training resource teachers to serve effectively students classified as mildly handicapped (Deno & Gross, 1973). My role was not only to develop the setting but also to act in the role of practicum supervisor, so I spent my days in a local elementary school working with the students and helping them to develop their intervention skills. An initial problem with which we were faced was how to decide what kind of intervention into a student's program was most appropriate. Although I had my own biases regarding the techniques students ought to use when they were attempting to improve a student's basic skills in an area like reading, I soon discovered that the practicum students had been imbued with a variety of different ideas from different faculty members in their dydactic coursework at the university. I wanted to take a dogmatic position that I as their practicum supervisor had the right to dictate the intervention procedures that they might use; unfortunately, as a scientist, I felt an obligation to remain open-minded regarding the alternatives proposed by my colleagues. After a period of uncertainty regarding how I should approach this task, I decided that the reasonable alternative was to address the problem empirically. The strategy I chose was to allow them to select any of a variety of alternative hunches that they might have regarding how a student might be taught, but to require that students evaluate the effects of whatever hunch they decided to try.

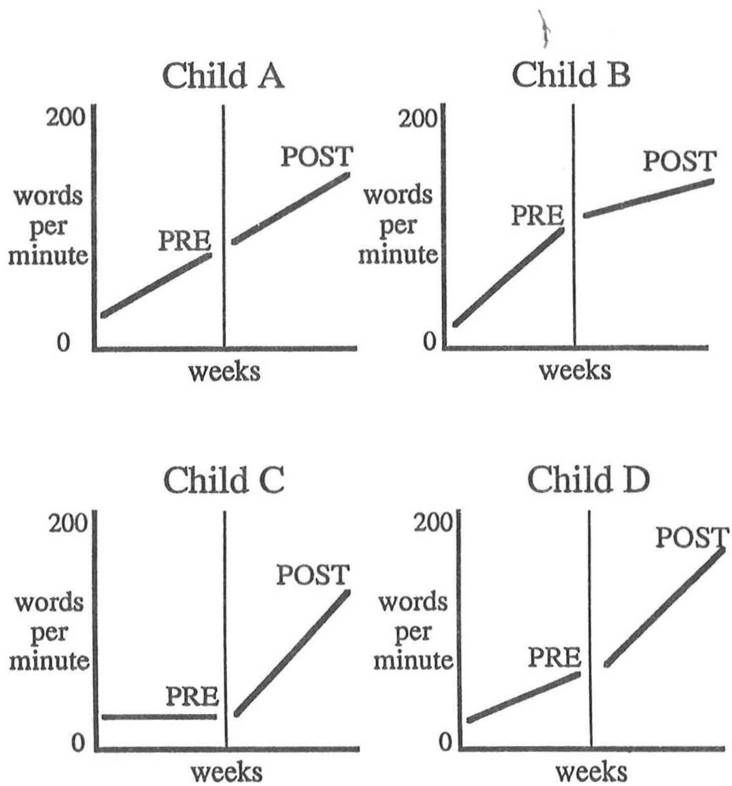
The problem with an open treatment and evaluation approach to making intervention decisions was, and is, how does one evaluate intervention effects with individual students? When teachers evaluate student growth at all, they typically do it on a posttest-only basis. Occasionally, in fields like special education, some effort is made to evaluate intervention effects by doing single-case pre/post

FIGURE 1
GRAPHIC DISPLAY OF A STUDENT'S READING SCORES



comparisions. As can be seen in Figure 2, however, even though growth may occur during the second phase (as shown by the straight ascending line between pre- and posttesting), our interpretation of that growth will differ, depending upon our knowledge of a child's growth rate prior to the intervention. For Child A, the pre-to-post growth rate is the same as that occurring prior to intervention. For Child B, the pre-to-post growth rate is actually lower than that which occurred prior to intervention. Only in the cases of Child C and D do we have evidence that the students' rate of growth increased in relation to intervention into the children's reading program.

Figure 2



The only feasible solution to the problem of evaluating the effects of interventions with individual students seems to be the use of single-case research design procedures. In single-use research designs, individual performance is measured repeatedly across time to produce a time-series data base that can be used for describing trends in student performance data under different intervention conditions. Thus, in the examples provided in Figure 2, the straight lines representing growth in student performance before and during intervention enable us to make comparative judgments regarding the conditions under which student growth occurs at a higher rate.

A real example of the use of repeated measurement of student performance across time to estimate slope differences in relation to intervention is shown in the data in Figure 3. These data were collected as part of a project to determine whether the effects of special education intervention could be evaluated using the single-case design model (Marston, 1988). As can be seen in Figure 3, both students increase in the rate they are acquiring reading fluency, beginning with the onset of special education. The effects of introducing special education for each student can, in this way, be evaluated, and the general effectiveness of special education can be estimated by aggregating individual cases. The basic schema represented in these two cases, then, provides us with a framework for considering the development of curriculum-based measurement.

Having made the decision to use single-case evaluation procedures to structure special educational interventions, our attention then turned to the development of an ongoing measurement system that teachers could use to establish the kind of data base necessary to produce the evaluation design presented in Figure 3. Since single-case designs require frequent repeated measurement, the question became both what to measure, and how to measure, student performance repeatedly to create the time-series data base required for single-case analysis.

Our initial efforts to develop measurement systems centered upon two approaches. The first approach was a rate of progress measure that was derived from data produced through monitoring the mastery of successive objectives in a sequence of skills or tasks across time (Deno & Mirkin, 1977). Mastery monitoring depends on criterion-referenced measurement of performance on specific tasks or skills typically laid out in a linear or hierarchical order. When using a rate of progress measurement system, the basic datum for evaluating intervention effects is change in the rate at which individual skills are mastered before and after intervention.

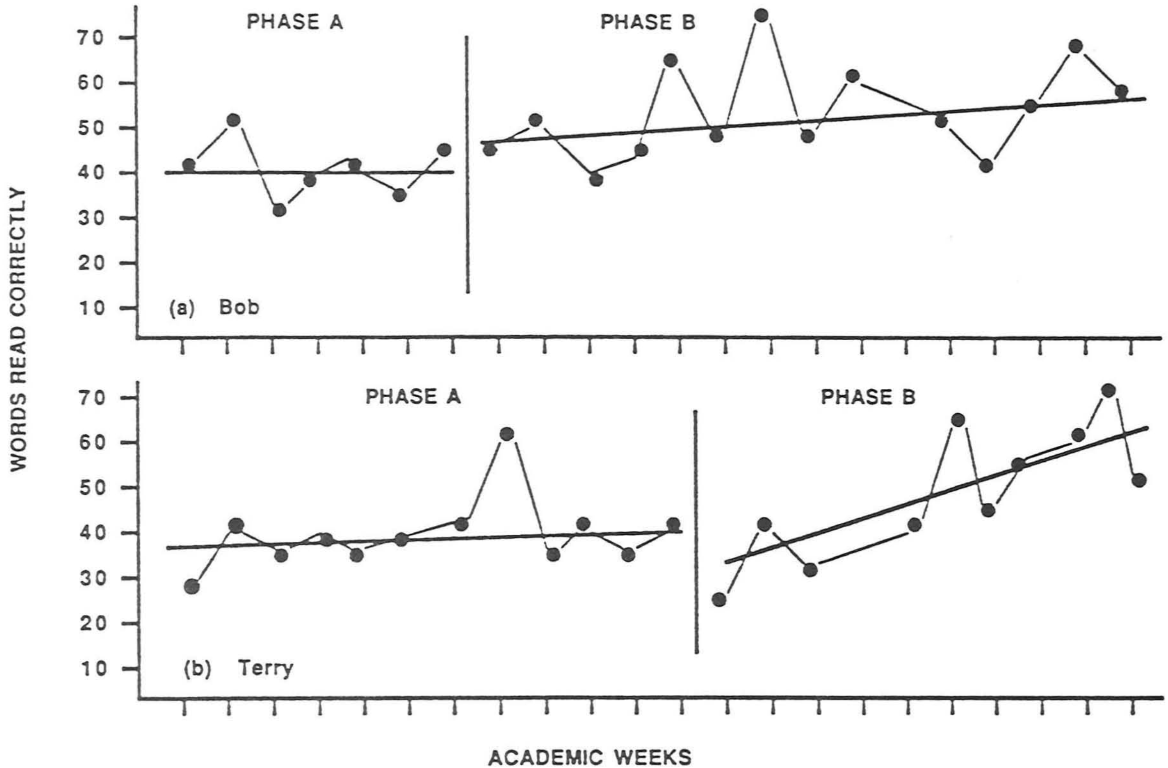


Figure 3. Multiple baseline performance of Bob and Terry

The second approach that we focused on was change in rate of performance on a single task, rather than rate of acquisition, or mastery, of multiple tasks (Deno & Mirkin, 1977). In contrast to the criterion-referenced mastery monitoring approach, this second approach involves specification of a single task on which repeated measurements can be obtained across a very long time period to describe change in proficiency on that task. A good example of measuring performance on a single task is the measurement of the amount of time taken to run a fixed distance, such as one mile. It is common for people who are interested in improving their endurance to monitor closely the amount of time that it takes them to run this fixed distance, and to use changes in the time taken to run the mile as a basis for making decisions about their training program. An analogous measurement system in education might be the length of time that it takes a very young child to print the letters of the alphabet. As a result of our research (Deno, 1985; Deno & Fuchs, 1987), we have come to favor the latter approach—measurement of change on a single task—for purposes of creating curriculum-based measurement procedures.

Reasons for Measuring Change on a Single Task

The rationale for favoring change in performance on an individual task, rather than mastery monitoring across multiple tasks, derives from several disadvantages of mastery monitoring and two advantages for measuring change in performance on a single task.

Mastery as a functional concept. The first problem or disadvantage with measuring the rate of progress in mastering tasks is that the technical and theoretical grounds of the approach are questionable. Three key assumptions must be true for mastery monitoring to be sensible. The first key assumption is that *mastery* as a construct is both theoretically and practically functional in the design and execution of instruction. The issues surrounding this assumption are complex and cannot be adequately considered here. However, the question that must be addressed is whether the acquisition of proficiency in the various curriculum domains actually occurs through mastery of discrete skills; and, following from that, whether instruction should be designed around subskill mastery. If so, then teaching to task mastery and monitoring progress in skill mastery is sensible. However, if student learning can proceed in many different ways for different students (i.e., learning is somewhat idiosyncratic), or if progress in the acquisition of proficiency can occur through partial mastery or skipping of various subskills, then a mastery learning model should not be reasonably imposed upon all students.

A second key assumption that must be met for a progress or mastery monitoring system to be sensible relates to the theoretical question just posed. If all students do not learn, or learn best, by meeting all the mastery criteria in a particular skill sequence, then does it make sense for all students to be required to meet the mastery criterion on each task within a skill sequence before moving on to a new learning experience? The significance of this consideration looms even larger when taking into account the fact that what constitutes mastery on a given task has rarely been empirically established and, therefore, that the mastery criterion specified for each task typically has been stipulated arbitrarily by the curriculum developer. Further, task sequences are almost always logically rather than empirically developed. Thus, the presumed transfer benefits obtained by requiring a student to achieve criterion performance on one task before moving to the next can only be speculations rather than assumptions. When considering these issues, it seems doubtful that teachers should pace their children on this basis. We need to be mindful that theoretical conceptions of children's learning and development ebb and flow, as evidenced by the current return to favoring more "wholistic" approaches. Mastery monitoring, as an assessment approach, is more typically assumed to be aligned with "reductionistic" models that rely on task analysis and isolated skill development. In contrast, CBM procedures function as global indicators of proficiency for different basic skills, and can be successfully used regardless of the particular theoretical conception of learning and cognition underlying curriculum and instructional design. For us, this has meant moving away from mastery monitoring systems that must be wedded to a particular approach to curriculum and instructional design.

A third key assumption that must be met for mastery monitoring systems to be sensible is that they be both technically and logistically feasible within the context of everyday instruction in the schools. The advent of microcomputers in the schools has made it possible to manage relatively complicated data sets in the classroom that can provide teachers with information on individual student progress. At present, however, the amount of information that teachers must process when monitoring individual student performance across several different basic skills exceeds practical limits. Further, as the number of subskills on which students are measured increases, the logistical problems increase for the teacher. Given advances in technology, this problem is not insurmountable; however, with CBM procedures we have tried to develop an approach that can be used in the current classroom without waiting for technological development.

Fractionation of learning. A second important problem associated with the mastery monitoring approach is that it fractionates the essential outcomes of learning in a particular curriculum domain. Thus, for the student, and often for the teacher, reading becomes performance on a series of isolated tasks represented as questions and answers on worksheets and curriculum-embedded mastery tests.

Too often, I am afraid, the result of this focus on isolated elements of the curriculum produces confusion in the minds of both teachers and students over the essential nature of what is being learned. Indeed, this overemphasis on the details of daily lessons is very likely what led Charles Silberman (1970), in an earlier call for educational reform, to identify "mindlessness" of educators regarding the purpose of education as the central problem of the schools. Students are affected, as well, by this fractionation of the curriculum. The dialogue between a special education teacher trainee, Diane, and her son, Ben, that is presented in Figure 4 illustrates what is most probably a common student viewpoint. The difference is that Ben is a very perceptive and articulate 7-year-old who seems to have reconciled the discrepancy between what his teacher does in the name of reading at school and what he has learned reading to be at home. In this dialogue, Ben makes clear that what he has learned to enjoy in the name of reading at home has very little to do with what he is required to do in the name of reading in school.

Figure 4

Conversation between Ben (7 years) and Mom, Fall '86

How come you always ask people about what they mean when they ask you if you like reading?

You know. Reading at home or reading at school.

Aren't they the same?

No. Like bat and bat.

Bat and bat?

Yeah. You know. A bat like a thing that flies in caves and a bat you hit a homerun with.

What does that have to do with reading at school or at home?

It's just the same. Reading and reading. You know.

No. I don't know. When you read at home you look at pages of a book, read the words, and find out from the words what happens. Isn't that the same as any reading? Isn't that the same thing reading is at school?

No. At school reading is looking at charts and doing worksheets and workbook and book and the teachers talks and stuff. You know. You're sposed to get them all right. Not fun.

Yes, but the book part, isn't that the same as reading other places?

No. You can't choose the stories and if you like it, it's not fun.

Why not, if you like the story you read.

Cause you can't finish it. If you do, you'll get in trouble.

Why? You're not supposed to go ahead of others?

Yeah. But I would get in trouble because if I sneaked and read the end, I wouldn't have time to finish my work.

D. Lilleberg

Although we might predict with some certainty that Ben will survive his school experience in reading, we may also speculate that the disinterest in reading exhibited by many secondary students and the shamefully high proportion of illiteracy among American adults has occurred because they became lost very early in the trees of their school's reading curriculum and never experienced the beauty of the forest that we know as reading.

Skill sequences as independent variables. The third problem we encountered when using mastery monitoring approaches was that we could not use the data generated through measuring student progress on the objectives to evaluate the use of alternative skill sequences. This problem occurs because, in mastery monitoring, the rate of progress on the skill sequence functions as the dependent variable. That is, mastery of the skills in the hierarchy defines the outcome, rather than the inputs, of instruction. If, we were interested in using the data generated through mastery monitoring to evaluate the use of a different curriculum that included a very different skill sequence, we could not do so. In effect, when one adopts a particular mastery monitoring system, one also adopts a particular scope and sequence of skills as the essential objectives of instruction. We wanted teachers to have data representing changes in dependent variables, independent from particular curriculum sequences, that could serve in evaluating alternative curricula and sequences of objectives. To do so required measurement procedures that were not wedded to one curriculum sequence.

Clarifying the focus. A fourth reason why we have opted for a measurement system based on measuring change in performance on a single task across time is that repeated measurements on the same task aids in focusing attention on an important proficiency indicator. This point, of course, is related to the "forest and the trees" problem, but the emphasis here is on the need for teachers to have clear and unambiguous feedback regarding the general effects of their instructional efforts. Too often, I think, teachers are either uncertain about the overall effects of their efforts to teach basic skills, or they are certain that they have been successful when a student has mastered the particular skills they have been teaching. In the first instance, their uncertainty stems from the fact that they have no "vital sign" indicators, such as pulse rate and temperature, that they can use to monitor the effects of their treatments on the educational health of their students. Indeed, I sometimes think teachers are like early flyers who had to resort to feel; that is, to "flying by the seat of their pants" because instruments to indicate aircraft altitude and attitude had not yet been developed. In the second instance, teachers' excessive certainty in their success stems from their overconfidence that specific skill mastery can be taken as evidence that the student is increasing in proficiency in the general curriculum where that skill is being taught. Since very little empirical justification ever exists for such an inference, the risk is real that teachers will conclude, as did the misguided surgeon, that "the surgery was a success, but the patient died."

Technical characteristics. Our final reason for building measurement procedures around change in performance on a single task was that the technical characteristics of such a system were superior to those of mastery measurement. Two facts, in particular, led to this conclusion. First, in our early efforts to assist teachers in developing and using progress measurement systems based on mastery measurement, we found the scores to be unreliable. This occurred because teachers tended to be inconsistent in their application of the mastery criteria—often for practical reasons, such as wanting to "keep the student up with his group," but sometimes for altruistic reasons, such as, "He came so close, I didn't want him to feel bad." Although each of us can appreciate why such reasons operate to produce variation from the mastery standard, it does not alter the fact that the data produced are of unknown reliability.

The second fact that led to our conclusion that scores based on repeated measurements of performance on a single task were preferable to those produced through mastery monitoring was that the scores produced by the former method were based on more nearly equivalent behavioral units than those produced when plotting progress in mastery of diverse skills. It is unreasonable to equate two separate reading subskills such as "identifying initial consonant blends" and "reading words with prefixes and suffixes" either behaviorally or cognitively. Any effort to plot graphically the mastery of these two tasks across time will most certainly reveal that students will take longer to master one than the other. When task or skill hierarchies are composed of a heterogeneous mix of skills of differing difficulty, it becomes virtually impossible to rely on a scale showing individual student progress in successively mastering those tasks, and to use graphs of student progress across time for evaluating the effects of changes in a student's instructional program. We believe that the actual performance scores obtained by repeatedly measuring student performance on the same task, using CBM procedures, are technically superior and more directly interpretable.

Selecting Tasks for Performance Measurement

Once we had established that our CBM procedures were to be based on repeatedly measuring performance on the same task, the importance of selecting the tasks for measurement became obvious. To stipulate and measure arbitrarily on tasks of unknown validity—so often the case when informal curriculum-based assessment occurs—would be indefensible.

A two-part strategy was used to identify those tasks that teachers might use in CBM. The first part of the strategy—initial task selection—was based on research using a criterion-validity paradigm to select those tasks that seemed to be the best candidates for repeated performance measurement (Deno, 1985). The second part of the task selection strategy was to test the tasks' instructional utility by evaluating the student achievement of teachers using the CBM data to make instructional evaluation decisions (Fuchs, Deno, & Mirkin, 1984; Fuchs, in press).

Criterion validity. In selecting tasks for the criterion-validity research, practical parameters were established for considering a task as a viable candidate for CBM:

1. *Repeatability.* Since the goal of measurement was to create a graphic time-series record of change in student performance, a task had to be one on which frequent repeated measurement could occur.
2. *Multiple forms.* Since repeated measurement was to occur and change in performance was to represent real growth in general proficiency, rather than the effects of practice on a specific task stimuli, a task had to be one for which it was simple to create many equivalent forms.
3. *Inexpensive.* Since many forms had to be made available for teachers to use frequently, the task had to be one that would not require costly materials.
4. *Time efficient.* Since frequent repeated measurement was required to create the graphic time-series record, the task needed to be one that did not consume too much instructional time.
5. *Easy to teach.* Since many teachers, paraprofessionals, and possibly students were to administer the measures, the task had to have one for which simple measurement procedures could be created and easily taught to nonprofessionals.
6. *Reliability.* Since the data were to be used to make important instructional intervention decisions, the tasks had to be ones for which reliable measures could be constructed.

Establishing parameters in task selection was important in the early program of research and development on CBM because it delimited the range and variety of tasks included in our search for valid indicators of reading proficiency. In addition, specifying the characteristics of a practically feasible task on which to do frequently repeated measurement enabled us to focus our criterion-validity research on only those tasks that could be part of a classroom-based, ongoing formative evaluation system.

The reasons for limiting task selection have not always been fully understood or appreciated by many, however. Indeed, the failure to include tasks for measurement that might operationally define the measurement domain more broadly is often mistakenly used as a basis for asserting that the CBM measures are invalid. A good illustration of the problem is in the area of reading, where we identified "reading aloud from text" as a task that can be used to create a global indicator of reading proficiency (Deno, Mirkin, & Chiang, 1982). The major criticism of measuring reading by having students read aloud from connected discourse is that such a task does not reflect a student's comprehension of text. On technical grounds, this criticism is invalid. The criterion validity research (summarized in Shinn, 1989) on using this task in reading measurement provides a solid empirical basis for concluding that the number of words read aloud correctly from text in a 1-minute time sample is a good indication of a student's general reading proficiency. CBM reading scores relate sensibly to standardized achievement test scores, to students' ages and grades, to teachers' judgments of reading proficiency, and to teachers' placements of students in regular, compensatory, and special education programs. Despite this, critics will argue that our CBMs in reading should include a "direct measure of comprehension," such as answering comprehension questions or retelling the story that has been read.

While it is possible to argue on empirical grounds that reading aloud from text indexes comprehension as well as so-called "direct measures" (cf. Fuchs, Fuchs, & Maxwell, 1988), it is more to the present point to clarify that tasks such as "answering comprehension questions" or "retelling the story" do not meet the requirements established for the measurement procedures we have been developing. To use either task would (a) consume far too much time to be used in a frequent measurement system (students would have to read fairly lengthy passages so that question asking or story retelling would be sensible); (b) cost too much in the development of multiple equivalent forms; and (c) in the case of story retell, be difficult to teach others to score reliably. Thus, although these tasks have been used as criterion measures in our validity research, they were excluded as candidates for our CBM procedures on other important grounds. We have painfully learned, however, that neither empirically nor technologically valid reasons are enough to persuade the critics. Clearly, face validity reigns supreme in education. A measure had better meet the consumer's preconceived notions of what an operational definition of the construct is supposed to look like if it is to be accepted easily. One cannot help but wonder if chemical engineers initially resisted the use of litmus paper because the "colors weren't right," or if doctors wouldn't use thermometers because they believed that a patient would "feel warm" if suffering from a fever. At the very least, we must conclude that, when it comes to measurement, educators are radically behavioral—operating as if inference beyond directly observed behavior is inappropriate.

Instructional utility. The criterion-validity data led us to conclude that it would be possible to teach teachers to use the CBM procedures to monitor routinely student performance and to evaluate the effects of daily instruction using the data thereby produced. Our hypothesis was that teachers using frequently collected data that graphically illustrated the rate of change in student performance could become more effective in timing their instructional change decisions, and that the result would be increased student achievement. To test this hypothesis, we designed a comparative study, in which special education teachers who used CBM in formatively evaluating their instruction were compared to teachers who used more conventional procedures (Fuchs, Deno, & Mirkin, 1984). The results of this study confirmed the hypothesis that teachers could increase students' achievement using the CBM procedures in formative evaluation. An important related outcome of the research was that evidence was obtained revealing that increases in CBM scores were related to increases in standardized achievement test scores, and most importantly, that increases in the number of words read aloud correctly in 1 minute across the school year were directly related to increases in the reading comprehension subtest scores of the students.

APPLICATIONS OF CBM

The results of the CBM research program have provided a basis for developing standardized measurement procedures that can be used to evaluate formatively the effects of modifications in the instructional programs for individual students. Indeed, the research conducted on the student achievement effects of special education teachers using these procedures provides a basis for concluding that instructional effectiveness can be improved through the use of CBM in formative evaluation (Fuchs, Deno & Mirkin, 1984; Fuchs, in press). At the same time, the CBM procedures have been used to "data-base" the full range of intervention decisions that are made for students who are academically at risk. These decisions include screening and monitoring high-risk students in the regular classroom program (Marston, 1988; Espin, Deno, Maruyama, & Cohen, 1989), evaluating prereferral interventions (Marston & Magnusson, 1985; Shinn, 1989), and developing IEPs (Deno, Mirkin, & Wesson, 1984), as well as reintegrating and follow-up monitoring of students terminated from special education services (Allen, 1989).

Advantages of CBM. Because traditional achievement measures have been used to "data-base" educational decisions for years, it is fair to wonder what the advantages of CBM might be. Several can be identified. First, because CBM data can be used to measure frequently performance across relatively short time periods, a new metric—slope—is available to evaluate interventions into individual student programs. The advantage of the slope metric is that it can be used to contrast the rate change in individual student performance under various instructional programs. Thus, teachers can execute a program, examine its effects on the rate of academic growth, change the program, examine the effects of the change relative to the previous program, and then decide whether to continue with the new program or to restore elements of the previous program. The continuous feedback regarding slope at various times enables teachers to make ongoing, data-based instructional decisions that are responsive to individual students. The net effect of using the slope data in this manner should be to improve cumulatively individual student programs. An illustration of the use of CBM data to improve cumulatively a student's program is presented in Figure 5. This figure is a graphic portrayal of the number of words read aloud correctly in 1 minute by Candy from his grade basal reader. Each heavy vertical line drawn on the graph identifies the point where a deliberate change was made by his teacher in an effort to find a more effective means of teaching him to read. The straight lines drawn through the data between vertical lines are a visual representation of the slope of Candy's performance during that phase of his program. As is evident from an overall inspection of Candy's progress, some of the changes introduced by his teacher into his program are associated with increases in slope and some are associated with decreases. Toward the end of the year, however, the overall trend in Candy's performance is increasing more rapidly than it was during the first half of the year. We cannot be certain that this more rapid rate of increase in performance is the result of his teacher's use of CBM data to continually evaluate his program and modify it in response to his performance; nevertheless, this is a plausible inference consistent with the research on the increased instructional effectiveness of teachers using CBM data in formative evaluation.

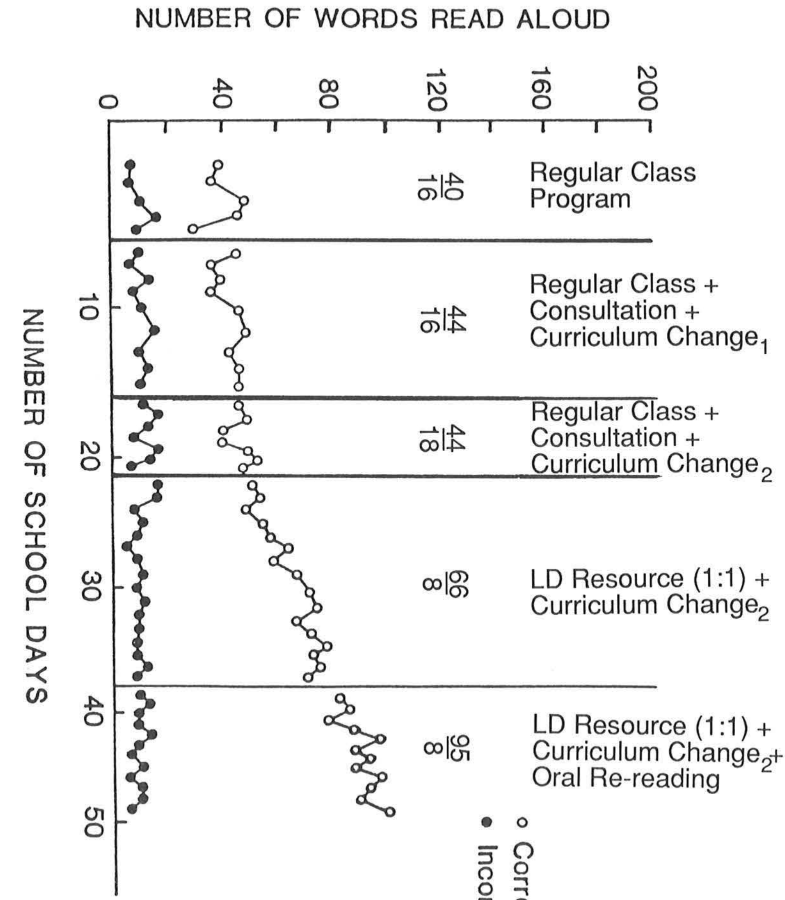


Figure 5

A second advantage of the CBM data is that they can more easily be used to communicate an individual student's progress in reading than is typically the case with commercially available standardized tests. This ease in communication derives from both the nature of the data presentation in CBM and the additional references available when CBM is set in the larger context of an ongoing evaluation system. The clarity of data presentation and interpretation is evident in viewing Figure 5. The number of words read correctly and incorrectly in 1 minute of reading from standard classroom text is not a datum that requires much explanation. Further, the simple line graph showing calendar dates and weekdays clearly reveals the level, trend, and variability of performance in student performance relative to significant periods of the school year. The utility of these graphs in communication was illustrated in the data collected by Fuchs, Deno, and Mirkin (1984). In that study comparing the effects of teachers using CBM data in systematic formative evaluation of individual students' programs, both the teachers and the students were able to specify correctly not only the students' IEP goals in reading, but also were able to predict accurately whether or not the students were going to make their goals. Comparison teachers using more conventional approaches to writing IEP goals and evaluating students' progress toward those goals could neither specify the goals at year's end, nor could they and their students correctly predict whether those goals would be attained. A strong argument can be made that a data system needs to be well and easily understood by those who are using it, if it is to become a functional part of students' programs.

CBM data graphs also communicate clearly because of the increased meaningfulness resulting from the increased number of references available when examining a student's graph. First, a student's performance is *curriculum referenced* in that the data reveal level, change, and variability in student performance on standard text material drawn from the student's local school and classroom. Second, a student's performance is *goal (or criterion) referenced* in that day-to-day performance can be compared both to the goal specified on the graph and to the daily increase required to attain that goal on the date specified for goal attainment. Third, a student's performance is *individually referenced* in that we can easily contrast the level, trend, and variability of the student's current performance with that same student's past performance. Fourth, student performance is *program referenced* in that it reveals how well the student progressed under different program arrangements or methods. Finally, a student's performance can be *norm referenced* by displaying how well a representative sample of that student's peers are doing in reading from the same material at the same time. A reading of Candy's graph in Figure 5 reveals all five types of references available in an individual student's CBM data graph. This rich array of referencing, easily and quickly apprehended in the graphic display of Candy's CBM data, becomes a powerful tool in the important communications surrounding an individual student's success in school.

Problems in implementing CBM. To describe CBM as if it is a measurement alternative with no associated problems or disadvantages would be misleading. In an effort to identify clearly the major barriers to implementing CBM, we conducted a Delphi survey of administrators and teachers who had implemented and were using CBM in their administrative units. The results of their inquiry are presented in Figures 6 and 7. A comparison of the administrators' and teachers' responses reveals a number of interesting differences. Teachers focus on the immediate impact of using CBM on a frequent basis and express concern about the additional time required in doing CBM. Three of the five most frequently identified barriers by teachers refer to time-associated problems. The remaining two teacher concerns relate to issues of measurement validity. As mentioned previously, the criterion-validity research rarely is powerfully persuasive with the teachers, and the face validity of CBM in reading and written expression is not high enough for many teachers. At the same time, less than 15% of the teachers who responded in the survey said they thought it was not a good idea that their district had implemented CBM.

The administrators' view of problems associated with implementing CBM was quite different from that of the teachers. The emphasis in the administrators' responses was that it was difficult to develop effective teacher use of the CBM procedures. Three of the five most frequently identified barriers by administrators addressed difficulties related to a lack of teachers' resourcefulness in using the CBM data responsively to modify and evaluate their instruction. Of interest is the fact that the single most frequently identified barrier from the administrators' perspective was the natural resistance that occurred when any change in practice was required of school personnel.

CONCLUSION

Curriculum-based measurement (CBM) has been presented here as an alternative to the more conventional measurement approaches available to educators—particularly special educators. Like curriculum-based assessment (CBA), CBM relies on direct observation of student performance on stimulus materials drawn directly from the local school curriculum. CBM is distinct from CBA in its specification of both what should be measured (i.e., the tasks) and how measurement should occur (i.e., the procedures). The gains accruing through the standardization used in CBM are those typical of improved technical adequacy in measurement: increased reliability and validity of the information obtained through measurement. Further, standardization

Figure 6

Delphi Probe - Teachers

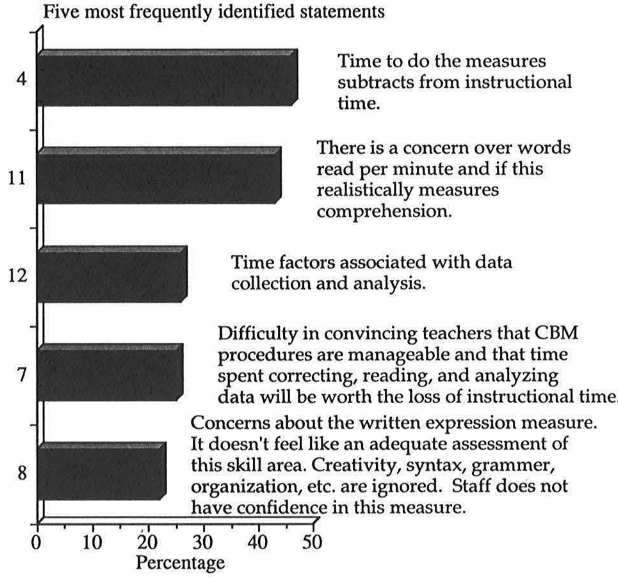
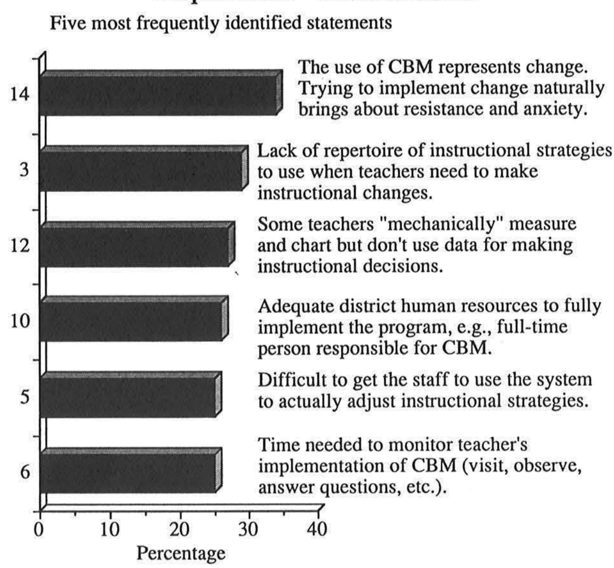


Figure 7

Delphi Probe - Administrators



permits aggregation of data across students for general program evaluation and establishes the conditions necessary for norm referencing. No gain is without loss, however. Standardization and prescription in measurement reduce the flexibility available through direct observation and recording of student behavior in the classroom and curriculum. The obvious solution to the problem of potential loss when using CBM is, of course, to train educators to use CBM and whatever other measurement procedures are appropriate in each individual case.

CBM has been developed to provide teachers with the tools to evaluate formatively the instruction they are providing to students who are developing functional literacy and numeracy. The goal has been to design procedures that teachers could use to make informed instructional decisions in such a way that they effect higher levels of achievement in their students than would otherwise be the case. Research evidence has accumulated that achievement increases can occur when teachers use CBM procedures to "data-base" their instruction. The research also makes clear that the connection between the simple collection of CBM data and increased achievement is not direct and automatic. The teacher's competence in using the data and designing alternative instruction mediates this relationship. When the CBM data signal the need for program change to a resourceful teacher, that teacher introduces program modifications that increase student success. The same signal sent to teachers who either are constrained by circumstances making change in students' programs impossible, or to teachers who do not know what else to do when a student is not learning, will not result in increased student achievement. There is no escaping from the fact that competent people are only made better when they use improved tools for doing their work and have the time and resources required for success.

REFERENCES

Allen, D. (1989). Periodic and annual reviews and decision to terminate special education services. In M. Shinn (Ed.), *Curriculum-based measurement*. New York: Guilford.

Bigge, J. (1988). *Curriculum-based instruction*. Mountain View, CA: Mayfield.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Deno, S., & Fuchs, L. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*, 1-16.

Deno, S., & Gross, J. (1973). The Seward-University Project: A cooperative effort to improve school services and university training. In E. Deno (Ed.), *Instructional alternatives for exceptional children*. Reston, VA: Council for Exceptional Children.

Deno, S., & Mirkin, P. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Deno, S., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.

Deno, S., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children, 16*, 99-104.

Espin, C., Deno, S., Maruyama, G., & Cohen, C. (1989). *The Basic Academic Skills Samples (BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms*. Unpublished manuscript available from S. Deno, Burton Hall, 178 Pillsbury Dr. S.E., Minneapolis, MN 55455.

Fuchs, L. (in press). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. Kramer (Ed.), *Curriculum-based assessment: Examining old problems, evaluating new solutions*. Hillsdale, NJ: Erlbaum.

Fuchs, L., Deno, S., & Mirkin, P. (1984). Effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.

Fuchs, L., Fuchs, D., & Maxwell, S. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20-28.

Howell, K., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education*. Columbus, OH: Merrill.

Idol, L., Nevin, A., & Paolucci-Whitcomb, P. (1986). *Models of curriculum-based assessment*. Rockville, MD: Aspen.

Marston, D. (1988). The effectiveness of special education: A time series analysis of reading performance in regular and special education settings. *Journal of Special Education, 21*(4), 13-26.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*(3), 266-276.

Shinn, M. (1989). *Curriculum-based measurement*. New York: Guilford.

Silberman, C. (1970). *Crisis in the classroom*. New York: Random House.

Tucker, J. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*(3), 199-204.